DOCUMENT RESUME

ED 163 079 TM 008 179

AUTHOR Meredith, John B., Jr.
TITLE Validation of a Simplified Method for Determining Passing Scores for Criterion-Referenced, Multiple-Choice Tests.
PUB DATE Mar 78
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Complexity Level; *Criterion Referenced Tests; *Cutting Scores; Item Analysis; *Mastery Tests; Multiple Choice Tests; *Reliability; *Scoring; *Teacher Role; Test Construction; Test Interpretation; Validity

ABSTRACT

The complexity of defining accurate passing scores with a minimum classification error when evaluating criterion-referenced, multiple-choice tests has been a major problem for classroom teachers. Therefore, a practical procedure in which the instructor determines the plausibility of each item option for the minimally acceptable examinee is proposed and examined for validity. A 20-item test was evaluated by subject-matter-experts (SMEs) to determine the Minimal Acceptable Performance Level (MAPL). An analysis of variance procedure was applied to the SME item-ratings, the results of which indicated no difference among SMEs. Further, it was found that the interjudgmental item-ratings were consistent. The test was administered to 14 Naval examinees whose mastery-nonmastery classifications based upon the MAPL were found to be in complete agreement with independent evaluations of the students by SMEs. It was concluded that this is a reasonable method which can be used easily by classroom teachers in a short amount of time.
(Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163079

Abstract

The purposes were to present and to validate an easily understood method for teachers' use in determining minimum passing scores for criterion-referenced evaluation of multiple-choice tests. Nedelsky (1954) originated this method employing subject-matter-experts (SMEs) ratings of test items. In this study, a 20-item test was evaluated by SMEs to determine the Minimal Acceptable Performance Level (MAPL). An analysis of variance procedure was applied to the SME item-ratings, the results of which indicated no difference among SMEs. Further, it was found that the interjudgmental item-ratings were consistent. This test was administered to 14 examinees whose mastery-nonmastery classifications based upon the MAPL were found to be in complete agreement with independent evaluations of the students by SMEs.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM008 179

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J B. Meredith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

VALIDATION OF A SIMPLIFIED METHOD FOR DETERMINING PASSING SCORES FOR CRITERION-REFERENCED, MULTIPLE-CHOICE TESTS

John B. Meredith, Jr.
Data-Design Laboratories
Norfolk, Virginia 23502

INTRODUCTION

When evaluating criterion-referenced, multiple-choice tests, the theoretical complexity of defining accurate passing scores with a minimum classification error presents a major problem to classroom instructors. The problem of determining passing scores is complicated by the large volume of relevant literature pertaining to criterion-referenced testing and the mathematical complexity of many of the approaches to the problem. Those complex approaches generally require the specification of classification errors, α and β, where the former is Type I error that falsely classifies a nonmaster as a master, and where the latter is a Type II error that falsley classifies a master as a nonmaster (Reichman and Oosterhoff, 1976). In most instances, instructors are unable to define these errors much less understand the trade-offs in setting appropriate α and β error levels.

This paper describes and validates the use of a practical procedure for classroom instructors that requires the instructor to determine the plausibility of each option to each test item for the minimally acceptable (just passing) examinee. The method was initially proposed by Nedelsky (1954) to establish an "absolute standard" for multiple-choice departmental physics course tests at the University of Chicago. In his review of methods for determining passing scores, Meskauskas (1976) found that only the Nedelsky procedure has received extensive practical use. The Nedelsky approach to the establishment of passing scores has been further refined and supported by Taylor and Reid (1972), Bobula (1974), Similansky and Guerin (1976), and Meredith (1977a).

In this study, the author digresses from the use of Minimal Acceptable Performance Level (MAPL) for training system assessment (Meredith, 1977a, b; Meredith and Robinson, 1976) to the use of the technique for determining the accuracy of passing scores with respect to mastery-nonmastery classifications for classroom instructors.

METHOD

The MAPL method for determining passing scores for a multiple-choice, criterion-referenced test is dependent upon the assumption that classroom instructors can define item alternative similarity as follows:

1. An alternative which a minimally acceptable examinee should recognize as incorrect is given a value of zero (0).

Printed in U.S.A.

2. An alternative which a minimally acceptable examinee should not recognize as incorrect is given a value of two (2).
3. An alternative which is correct is given a value of two (2).
4. All other alternatives are given values of one (1).

The application of the MAPL technique is exemplified by the classroom instructor who is teaching a first-year algebra course and defines the ability to recognize the identity property of multiplication as a component of basic algebraic competency. The instructor might write an item and assign the following alternative similarity values:

Value	Item
	"The identity property of multiplication is..."
2	1. a x 0 = 0
2 (correct)	2. a x 1 = a
1	3. a + 0 = a
0	4. a - a = 0

When the instructor evaluates the item for the minimally acceptable examinee, he decides that the fourth alternative does not meet the minimal algebraic competency for the identity property of multiplecation, and assigns it a value of zero. The instructor decides that the first alternative is a viable distractor and assigns it a value of two. The second alternative is the correct response and the instructor assigns it a value of two. The instructor assigns a value of one to the remaining alternative because it is a marginal alternative that contains the identity property of addition.

The Alternative Similarity Index (ASI) for the preceeding example is defined as follows:

$$ASI = \frac{2}{2 + 2 + 1 + 0} = \frac{2}{5} = .4$$

where the numerator is the value assigned to the correct response, (2), and the denominator is the sum of all values assigned to the item (5).

For implementation of the MAPL technique, more than one classroom instructor should be utilized for determination of each item ASI. If a sufficient number of classroom instructors is not available, other Subject Matter Experts (SMEs) may be utilized for determining each item ASI.

A generalized formula for calculating each ASI is defined as follows:

$$ASI_i = \frac{2n}{\sum_{j=1}^{n} \sum_{k=1}^{m} A} \qquad (1)$$

where:

i = item number,
n = number of SMEs,
m = number of alternatives, and
A = value assigned to m^{th} alternative by the n^{th} SME.

The MAPL (minimum passing score) of a test is determined by summation of the ASIs. Any examinee whose number-right test score meets or exceeds the MAPL is deemed a master. Conversely, any examinee whose-right number test score does not meet the MAPL is deemed a nonmaster.

The validation of the MAPL technique in this study was approached from two different directions. First, validation of SME item-ratings was accomplished by testing the null hypothesis of no difference among SME ratings, using a one-way analysis of variance design, and by determining the interjudgmental reliability of the SMEs item-ratings. Second, actual mastery-nonmastery classifications of examinees were validated by SMEs.

FINDINGS

In the study, five classroom instructors evaluated a 20-item test to determine the ASI for each item. From examination of the 5 x 20 matrix of ASIs, it was ascertained that some of the ASIs were obvious outliers. Six of the ASIs were trimmed from the ASI matrix and were not utilized in the analysis. However, no more than one observation per item was deleted by an outlier criterion based upon the mean and standard deviation of the item ASI rating. (Dixon and Massey, 1969).

A one-way analysis of variance design was utilized to test for differences among the five classroom instructors. The null hypothesis was:

$$Ho: \quad M_a = M_b = M_c = M_d = M_e$$

An F-ratio of 1.3122 with 4 and 89 degrees of freedom was found not to be significant at the .05 level. Therefore, the null hypothesis of no difference among classroom instructor item ASI ratings was not rejected.

In order to determine the interjudgmental reliability, a repeated measure analysis of variance procedure was utilized to determine the ratio of estimated true score variance to the sum of estimated true score variance and estimated error score variance (Winer, 1971). The interjudgmental reliability, when utilizing five SMEs, was .78.

This 20-item test was administered to 14 Navy submarine technicians (N = 14) who were attending a replacement training course. The MAPL, sum of item ASI ratings, was found to be 10.3. The number-right score range on the test was from 10 to 17. When comparing the examinee scores to the MAPL of 10.3, only two examinees, who had

scores of 10, were found to be nonmasters. (Those two examinees were academically dropped from the replacement training program because of substandard performances in other courses.) The five classroom instructors of the course independently evaluated all 14 examinees and found that only those two examinees were not minimally acceptable with regard to actual equipment responsibilities.

CONCLUSIONS

This paper has presented a method that can be easily utilized by classroom instructors for evaluating multiple-choice, criterion-referenced tests. The method was validated by an analysis of variance design testing inter-instructor item ratings and testing intra-instructor rating consistency and by the criterion validity of examinee mastery-nonmastery classifications.

It was found that there were no significant differences among classroom instructor item ratings, while maintaining a high reliability index (.78), of judgmental ratings. It was also found that classroom instructors identified the same examinees as being nonmasters as had the MAPL procedure, thus confirming the criterion validity of examinee classifications by classroom instructors.

The educational implication of the MAPL procedure is that it provides a means for classroom instructors to establish reasonable passing scores without committing a disproportionately large amount of their time. The ease with which the method, compared to other methods for establishing passing scores, can be used for competency-based educational programs is essential for classroom instructors. After all, classroom instructors are a very valuable limited resource; and they need efficient methods for criterion-referenced evaluation.

REFERENCES

1. Bobula, J. Minimum Pass Level Study. Report to the Faculty, Center for Educational Development, College of Medicine, University of Illinois, 25-31, 1974.

2. Dixon, W. J. and Massey F. J., Jr. Introduction To Statistical Analysis, New York, McGraw-Hill, 1969.

3. Meredith, J. B., Jr. The Use of A Criterion-Referenced Test Methodology to Assess Training System Effectiveness, presented at The American Educational Research Association, New York, New York, April, 1977.

4. Meredith, J. B., Jr. The Evaluation of Training System Effectiveness By The Use of A Criterion-Referenced, Multiple-Choice Test Methodology, In Press, 1977.

5. Meredith, J. B., Jr. and Robinson, M. A. Determination of Minimal Acceptable Performance Levels For Criterion-Referenced, Multiple-Choice Tests, Proceedings of the 18th annual meeting of the Military Testing Association, 1976.

6. Meskauskas, J. A. Evaluation Models for Criterion-Referenced Testing: Views Regarding Mastery and Standard Setting, Review of Educational Research, 46:133-158, 1976.

7. Nedelsky, L. Absolute Grading Standards for Objective Tests. Educational and Psychological Measurement, 14:3-19, 1954.

8. Reichman, S. L. and Oosterhoff, A. C. Strategy Guidelines for the Construction of Mastery Tests, Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, April, 1976.

9. Smilansky, J. and Guerin, R. O. Minimal Acceptable Performance Levels for Criterion-Referenced Multiple-Choice Examinations and Their Validation, Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, California, April, 1976.

10. Taylor, D. D. and Reid, J. C. Criterion-Referenced Evaluation of Student Performance, Journal of Medical Education, 47:970-971, 1972.

11. Winer, B. J. Statistical Principles of Experimental Design, New York, McGraw-Hill, 1971.